9Ш

УДК 004.934

М.М. Шатковський
Інститут кібернетики ім. В.М. Глушкова НАН України, м. Київ, Україна
mykolash@gmail.com

Об’єктно-елементна модель подання інформації для задачі конкатенативного сегментивного синтезу української мови

У статті наводиться модель подання інформації для задачі конкатенативного сегментивного синтезу української мови. Демонструється підхід у виборі об’єктів та елементів синтезу. Аналізується структура об’єктів та елементів синтезу бази даних системи конкатенативного сегментивного синтезу української мови.

Вступ

Синтез природних мов є важливим функціональним компонентом систем штучного інтелекту, оскільки підвищує інтелектуалізацію людино-комп’ютерного інтерфейсу та надає доступ до природного для людини способу спілкування.

Метою сучасних систем синтезу природної мови є озвучення інформації з якнайприроднішим наближенням характеристик звучання до голосу людини.

На сьогоднішній день в Україні активно досліджують та успішно вирішують проблеми синтезу та розпізнавання голосових мовних сигналів в Українській асоціації з оброблення інформації та розпізнавання образів при МННЦІТС [1], в Державному університеті інформатики і штучного інтелекту [2], в Інституті кібернетики імені В.М. Глушкова НАН України [3], в Київському національному університеті імені Тараса Шевченка [4] та ін.

Для вирішення задачі побудови систем природного синтезу мови слід опрацювати ряд наукових проблем, серед котрих особливо важливими є такі – дослідження сприйняття слухачем характеристик звучання синтезованих мовних сигналів та адекватне розуміння озвученої інформації; обробка, подання та інтерпретація вхідної інформації для озвучення; розробка множин елементів синтезу та характеристик їх звучання та ін.

Важливою підзадачею створення систем природного синтезу мови є розробка принаймні таких програмних засобів, як засоби обробки та налаштування вхідної текстової та голосової інформації для створення вибірок мовної інформації згідно з моделлю синтезу мови; засоби розробки, зберігання та використання баз даних елементів синтезу; засоби озвучення обробленої вхідної текстової інформації, використовуючи бази даних елементів синтезу.

Методи конкатенативного синтезу мови в сучасних системах озвучення текстової інформації

При виборі елементів синтезу необхідно враховувати велику кількість фонетичних коартикуляційних властивостей елементів мови, зокрема, таких як:
– акустичні особливості і характеристики ізольованих звуків та звуків у слові;
– просодичні характеристики та інтонаційні властивості природної мови в залежності від їх розташування в слові, а також їх пом’якшення, подвоєння й ін.;

– вплив акцентованості і наголосу на частотні характеристики мови;
– суперсегментні явища та характеристики частоти основного тону;
– структуризацію пари «об’єкт синтезу – елемент синтезу».

Встановлено, що фізіологічний тремор м’язів гортані створює невеликі випадкові зміни їх напруження, які впливають на сприйняття мови [5].

В залежності від елементів синтезу, що використовуються при генеруванні голосових мовних сигналів в сучасних системах озвучення текстової інформації, виділяються такі методи конкатенативного синтезу.

1. Алофонний конкатенативний синтез.

У мовному потоці фонеми реалізуються як алофони, іншими словами, як позиційні та комбінаторні відтінки фонем. Позиційний фактор враховує позицію даної фонеми відносно мовленнєвого, акцентно-групового, синтагматичного та фразового наголосу. Комбінаторний фактор враховує найближче фонемне оточення [6].

2. Дифонний конкатенативний синтез.

Елементами синтезу є звукові мовні одиниці, що містять половину приголосного та половину голосного звуку, що примикає до нього. Саме коартикуляційні переходи між фонемами містять переважну більшість семантичного навантаження в слові.

3. Складовий конкатенативний синтез.

Елементами синтезу зазвичай є склади (частіше лише відкриті) та фонеми. Можливе використання як з’єднуючих мовних елементів алофонів та дифонів.

4. Трифонний.

Трифони – мовні фонетичні одиниці, що починаються з середини стаціонарної частини вокалізованого звуку, містять перехідні частини з першого звуку до другого та з другого до третього, всю стаціонарну частину другого звуку та закінчуються серединою стаціонарної частини після наступного звуку [7].

5. Сегментивний – синтез із одиниць конкатенації довільного розміру.

Слово розкладається на три частини – початкову (префіксну), внутрішню та кінцеву (суфіксну). В множинах сегментів містяться як наголошені сегменти, так і ненаголошені. А слово подається послідовністю певних сегментів, які вже містять в собі просодику слова. Самі ж сегменти вибираються (із відповідної множини сегментів) якнайбільшими, тобто такими, що складаються з якнайбільшої кількості фонем. Це зменшує кількість конкатенацій сегментів та підвищує природність звучання синтезованої мови.

Розглянемо математичну модель подання текстової інформації для конкатенативного сегментивного синтезу української мови, запропоновану в (1):

$$w_l = s_j \{i\} f_k , \tag{1}$$

де $w_l \in W, \forall l \in N$ – слово із простору слів української мови W, $s_j \in S, \forall j \in N$ – деякий початковий сегмент із множини початкових сегментів S, $\{i\}$ – деякий (певний для кожного конкретного слова) послідовний набір внутрішніх сегментів $i_m \in I, \forall m \in N$ із множини внутрішніх сегментів I, $f_k \in F, \forall k \in N$ – деякий кінцевий сегмент із множини кінцевих сегментів F, $s_j, i_m, f_k \in \Omega, \forall j, m, k \in N$, де Ω – простір всіх сегментів української мови.

Наведена модель дозволяє істотно підвищити швидкість роботи систем звукового синтезу мовної інформації завдяки відносній простоті концепції та зручності обробки і доступності елементів синтезу [8].

На рис. 1 зображено врахування властивостей звучання різних мовних елементів

для об’єкта синтезу відповідно до моделі (1):

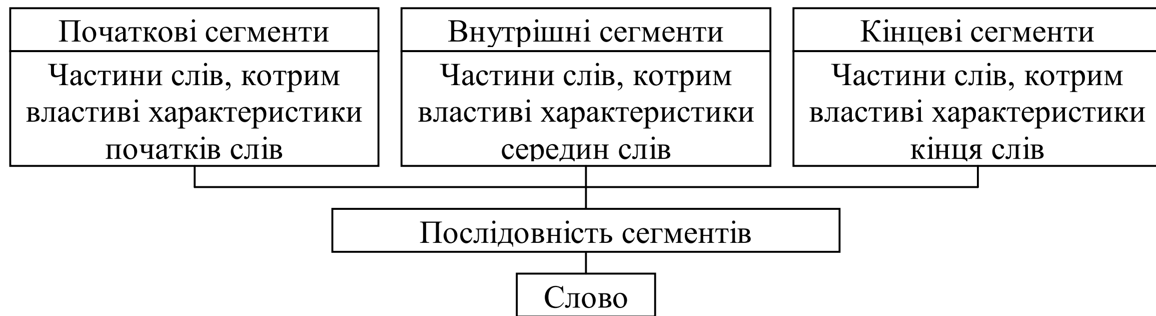


Рисунок 1 – Схема слова як об’єкта синтезу для конкатенативного сегментивного синтезу української мови

Такі одиниці синтезу вибираються для того, щоб уникнути складності штучного моделювання процесів коартикуляції завдяки адекватному відтворенню просодичних властивостей природних мовних сигналів.

Основні наукові зусилля з області розробки систем синтезу природної мови без модифікації елементів мови спрямовані на вдосконалення баз елементів синтезу та алгоритмів сегментації для покращення характеристик природності звучання та просодичного оформлення синтезованої мови.

Частина елементів синтезу може володіти спільними для декількох мов фонемними та просодичними характеристиками, але розробляти мультимовні системи синтезу зі звучанням мови аналогічним природному можна лише при використанні сучасних математичних моделей та алгоритмів модифікації вхідних елементів синтезу. Наприклад, засобами модифікації періодів основного тону, так званими PSOLA-методами [9].

Об’єктно-елементна модель подання інформації для конкатенативного сегментивного синтезу української мови

Роль слова як центральної одиниці мови зумовлюється залежністю інших одиниць мови від слова та системними відношеннями зі словом. Дедалі більше лінгвістів визнають правильним твердження про те, що саме слово є основною одиницею мови [10].

Часто в розмовній мові послідовність фонем, що відповідає слову, не співпадає з послідовністю літер, відповідних фонемам. Тобто, відрізняється від буквальної конкатенації фонем. Це явище називається несвідомою фонетизацією і відбувається в силу ряду причин, в тому числі і фонетично обумовлених:

– дзвінкі і глухі приголосні, взаємодіючи у мовному потоці, можуть уподібнюватися один одному, що відповідно відображається у вимові;
– приголосні, в залежності від оточуючих фонем, можуть вимовлятися м’яко;
– при збігові кількох приголосних у вимові один із них випадає, тобто відбувається спрощення приголосних;
– приголосні можуть подовжуватися;
– голосні, в залежності від наголошеності та оточуючих їх інших голосних, можуть вимовлятися як чітко і «неспотвореними», так і, в силу вокалізованої природи голосних звуків, з наближенням до інших голосних.

В описаній задачі мінімальним семантично повним носієм інформації є слово. Це підказують і результати фонетичних досліджень, які показали принципову залежність просодичних характеристик звучання мови від їх розташування в слові [11].

Розглянемо об’єктно-елементну модель подання інформації, наведену на рис. 2.

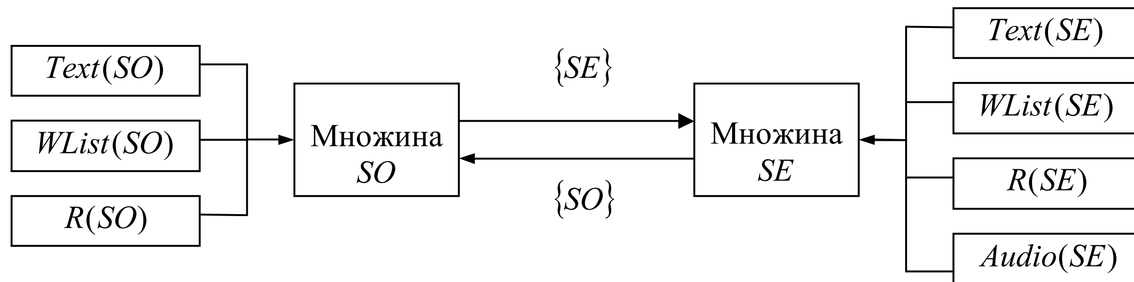


Рисунок 2 – Схема об’єктно-елементної моделі подання інформації

Об’єкт синтезу (слово) подається послідовною множиною елементів синтезу $\{SE\}$. Кожний елемент синтезу SE має таку структуру – текстове подання елементу синтезу $Text(SE) \in \Omega$; список слів, в яких міститься елемент синтезу $WList(SE) = \{w_i\}$; де $i \in N, w \in W$; рейтинг, тобто кількість слів, в яких міститься елемент синтезу $R(SE)$, звукове подання елементу синтезу $Audio(SE)$. Кожний об’єкт синтезу SO має таку структуру – текстове подання об’єкта синтезу $Text(SO) \in W$; послідовний список елементів синтезу, які складають слово $WList(SO) = s_j \{i\} f_k$, $s_j \in S, \forall j \in N$, $\{i : i_m \in I, \forall m \in N\}$, $f_k \in F, \forall k \in N$, $s_j, i_m, f_k \in \Omega, \forall j, m, k \in N$; рейтинг, тобто кількість елементів синтезу, які складають слово $R(SE)$.

Структура об’єктів та елементів синтезу бази даних системи конкатенативного сегментивного синтезу української мови

Конкатенативні синтезатори залежать від мовних баз даних. Чим більший об’єм мовної бази даних, тобто чим повніше подана в ній звукова, інтонаційна, темпоральна варіативність мови, тим більшу природність звучання можна отримати в синтезованій мові [12].

Наведемо спрощений приклад аналізу структури об’єктів синтезу (табл. 1) та елементів синтезу (табл. 2) бази даних системи конкатенативного сегментивного синтезу української мови на прикладі фрази «Об’єктно-елементна модель подання інформації для конкатенативного сегментивного синтезу української мови».

Таблиця 1 – Структура об’єктів синтезу бази даних системи конкатенативного сегментивного синтезу української мови

Об’єкт синтезу	Впорядкований список сегментів	К-сть сегментів	К-сть літер
об’єктно	#о,б’є_, к, т, но^	5	7
елементна	#е, ле, ме_нт, на^	4	9
модель	#мо, де_, ль^	3	6
подання	#по, да, ння_^	3	7
інформації	#ін, фо, р, ма_, ції^	5	10
для	#для_^	1	3
конкатенативного	#ко, н, ка, те, на, ти_, в, ного^	8	16
сегментивного	#се, г, ме, н, ти_, в, ного^	7	13
синтезу	#си_, н, те, зу^	4	7
української	#у, к, ра, ї_, н, ської^	6	11
мови	#мо_, ви^	2	4

Основною, для конкатенативного сегментивного синтезу, характеристикою об'єкта синтезу є впорядкована послідовність його елементів синтезу. Адже основні мовні властивості та особливості звучання, якими оперує метод конкатенативного сегментивного синтезу, вже закладені саме в елементах синтезу згідно з парадигмою створення їх множини.

Таблиця 2 – Структура елементів синтезу бази даних системи конкатенативного сегментивного синтезу української мови

№	Елемент синтезу	Список об'єктів синтезу	К-сть об'єктів	К-сть літер	№	Елемент синтезу	Список об'єктів синтезу	К-сть об'єктів	К-сть літер
1	#для_^	для	1	3	20	зу^	синтезу	1	2
2	#ської^	української	1	5	21	ви^	мови	1	2
3	ме_нт	елементна	1	4	22	#о	об'єктно	1	1
4	ння_^	подання	1	3	23	#е	елементна	1	1
5	ного^	конкатена-тивного сегментив-ного	2	4	24	ле	елементна	1	2
6	б'є_	об'єктно	1	2	25	да	подання	1	2
7	ції^	інформації	1	3	26	фо	інформації	1	2
8	#си_	синтезу	1	2	27	ка	конкатена-тивного	1	2
9	#мо_	мови	1	2	28	те	конкатена-тивного синтезу	2	2
10	на^	елементна	2	2	29	на	конкатена-тивного	1	2
11	но^	об'єктно	1	2	30	ме	сегментив-ного	1	2
12	#мо	модель	1	2	31	#у	української	1	1
13	де_	модель	1	2	32	ра	української	1	2
14	ль^	модель	1	2	33	ї_	української	1	1
15	#по	подання	1	2	34	к	об'єктно української	2	1
16	#ін	інформації	1	2	35	т	об'єктно	1	1
17	ма_	інформації	1	2	36	р	інформації	1	1
18	#ко	конкатена-тивного	1	2	37	н	конкатена-тивного сегментив-ного синтезу української	4	1
19	ти_	конкатена-тивного сегментив-ного	2	2					

З табл. 3 видно, що більшість сегментів одночасно враховує декілька характеристик природності звучання мовних одиниць.

Таблиця 3 – Аналіз характеристик елементів синтезу бази даних системи конкатенативного сегментивного синтезу української мови

№	Характеристики природності звучання сегментів	Сегменти
1	Наголошеність голосних (в усіх наголошених сегментах)	б'є_, ме_нт, де_, ння_^, ма_, #для_^, ти_, #си_, ї_, #мо_
2	Врахування природних переходів від приголосних до голосних	в усіх сегментах з приголосними
3	Просодичні характеристики звуків на початку слів	#о, #е, #мо, #по, #ін, #для_^, #ко, #се, #си_, #у, #мо_
4	Просодичні характеристики звуків в середині слів	б'є_, к, т, ле, ме_нт, де_, да, фо, р, ма_, н, ка, те, на, г, ме, ти_, в, ра, ї_,
5	Просодичні характеристики звуків в кінці слів	но^, на^, ль^, ння_^, ції^, #для_^, ного^, зу^, ської^,ви^
6	Врахування характерних властивостей пом'якшеності приголосних	б'є_, ль^, ської^
7	Врахування характерних властивостей звучання морфологічних конструкцій	ме_нт
8	Врахування характерних властивостей звучання подвоєння приголосних	ння_^
9	Врахування характерних властивостей взаємовпливу приголосних звуків	ської^, ме_нт, ння_^

Висновки

Розглядається модифікований підхід до вибору елементів синтезу української мови. Новизна підходу полягає у використанні запропонованої у статті об'єктно-елементної моделі синтезу для створення зв'язків між об'єктами синтезу (словами) та елементами синтезу. Описана модель обґрунтовує використання слова як центральної одиниці інформації, оскільки слово є мінімальним семантично повним носієм інформації та залежить від фонетичних коартикуляційних властивостей елементів мови. Наведена модель дозволяє уникнути складності штучного моделювання процесів коартикуляції завдяки відтворенню просодичних властивостей природних мовних сигналів. На тестовому прикладі розглянуто використання об'єктно-елементної моделі подання інформації для конкатенативного сегментивного синтезу української мови.

Подальші дослідження будуть направлені на створення множин та мовних баз даних (текстових і звукових) об'єктів та елементів синтезу (сегментів) і текстового корпусу для задачі конкатенативного сегментивного синтезу української мови.

Література

1. Режим доступу: http://uasoiro.org.ua
2. Режим доступу: http://iai.donetsk.ua
3. Режим доступу: http://icyb.kiev.ua

4. Режим доступу: http://univ.kiev.ua/
5. Сорокин В.Н. Теория речеобразования. – М.: Радио и связь, 1985. – 312 с.
6. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. – Изд-во: «Белорусская наука», 2008. – 342 с.
7. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К: Наук. думка, 1987 – 264 с.
8. Крак Ю.В., Шатковський М.М. Математична модель подання текстової інформації для методу конкатенативного сегментивного Text-to-Speech синтезу // Штучний інтелект. – 2006. – № 1. – С.158-165.
9. T. Dutoit and H. Leich. MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database // Speech Communication, 1993.
10. Вихованець І.Р. Частини мови в семантико-граматичному аспекті. – К., 1988 – С. 14-19.
11. Кривонос Ю.Г., Крак Ю.В., Шатковский Н.Н. Структура, свойства, характеристики объектов и элементов синтеза речи // Компьютерная математика. – 2006. – № 1. – С. 61-69.
12. Людовик Т.В., Сажок Н.Н. Использование речевых баз данных большого объема при синтезе речи в системах искусственного интеллекта // Проблемы управления и информатики. – 2003. – № 6. – С. 82-87.

Н.Н. Шатковский
Объектно-элементная модель подачи информации для задачи конкатенативного сегментивного синтеза украинской речи
В статье приводится модель подачи информации для задачи конкатенативного сегментивного синтеза украинской речи. Демонстрируется подход в выборе объектов и элементов синтеза. Анализируется структура объектов и элементов синтеза базы данных системы конкатенативного сегментивного синтеза украинской речи.

M. Shatkovskyy
Object-elemental Model of Text Information Representation for Concatenation Segmentation Synthesis
Information representation model for segmentation concatenation Ukrainian synthesis is applied. Synthesis objects and elements selection approach is demonstrated. Synthesis objects and elements structure for database of segmentation concatenation Ukrainian synthesis system is analyzed.

Стаття надійшла до редакції 30.07.2008.